Maximum likelihood estimators for the extreme value index based on the block maxima method

Clément Dombry*

January 24, 2013

Abstract

The maximum likelihood method offers a standard way to estimate the three parameters of a generalized extreme value (GEV) distribution. Combined with the block maxima method, it is often used in practice to assess the extreme value index and normalization constants of a distribution satisfying a first order extreme value condition, assuming implicitely that the block maxima are exactly GEV distributed. This is unsatisfactory since the GEV distribution is a good approximation of the block maxima distribution only for blocks of large size. The purpose of this paper is to provide a theoretical basis for this methodology. Under a first order extreme value condition only, we prove the existence and consistency of the maximum likelihood estimators for the extreme value index and normalization constants within the framework of the block maxima method.

Key words: extreme value index, maximum likelihood estimator, block maxima method, consistency.
AMS Subject classification: 62G32.

1 Introduction and results

Estimation of the extreme value index is a central problem in extreme value theory. A variety of estimators are available in the literature,

*Université de Poitiers, Laboratoire de Mathématiques et Applications, UMR CNRS 7348, Téléport 2, BP 30179, F-86962 Futuroscope-Chasseneuil cedex, France. Email: clement.dombry@math.univ-poitiers.fr

for example among others, the Hill estimator [8], the Pickand's estimator [13], the probability weighted moment estimator introduced by Hosking *et al.* [10] or the moment estimator suggested by Dekkers *et al.* [5]. The monographs by Embrechts *et al.* [7], Beirlant *et al.* [2] or de Haan and Ferreira [4] provide good reviews on this estimation problem.

In this paper, we are interested on estimators based on the maximum likelihood method. Two different types of maximum likelihood estimators (MLEs) have been introduced, based on the peak over threshold method and block maxima method respectively. The peak over threshold method relies on the fact that, under the extreme value condition, exceedances over high threshold converge to a generalized Pareto distribution (GPD) (see Balkema and de Haan [1]). A MLE within the GPD model has been proposed by Smith [18]. Its theoretical properties under the extreme value condition are quite difficult to analyze due to the absence of an explicit expression of the likelihood equations: existence and consistency have been proven by Zhou [21], asymptotic normality by Drees *et al.* [6]. The block maxima method relies on the approximation of the maxima distribution by a generalized extreme value (GEV) distribution. Computational issues for ML estimation within the GEV model have been considered by Prescott and Walden [14, 15], Hosking [9] and Macleod [11]. Since the support of the GEV distribution depends on the unknown extreme value index γ, the usual regularity conditions ensuring good asymptotic properties are not satisfied. This problem is studied by Smith [17]: asymptotic normality is proven for $\gamma > -1/2$ and consistency for $\gamma > -1$.

It should be stressed that the block maxima method is based on the assumption that the observations come from a distribution satisfying the extreme value condition so that the maximum of a large numbers of observations follows approximatively a generalized extreme value (GEV) distribution. On the contrary, the properties of the maximum likelihood relies implicitely on the assumption that the block maxima have *exactly* a GEV distribution. In many situations, this strong assumption is unsatisfactory and we shall only suppose that the underlying distribution is in the domain of attraction of an extreme value distribution. This is the purpose of the present paper to justify the maximum likelihood method for the block maxima method under an extreme value condition only.

We first recall some basic notions of univariate extreme value theory. The extreme value distribution distribution with index γ is noted G_γ and has distribution function

$$F_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \quad 1+\gamma x > 0.$$

We say that a distribution function F satisfies the extreme value condition with index γ, or equivalently that F belongs to the domain of attraction of G_γ if there exist constants $a_m > 0$ and b_m such that

$$\lim_{m\to+\infty} F^m(a_m x + b_m) = F_\gamma(x), \quad x \in \mathbb{R}. \tag{1}$$

That is commonly denoted $F \in D(G_\gamma)$. The necessary and sufficient conditions for $F \in D(G_\gamma)$ can be presented in different ways, see e.g. de Haan [3] or de Haan and Ferreira [4, chapter 1]. We remind the following simple criterion and choice of normalization constants.

Theorem 1. *Let $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$ be the left continuous inverse function of $1/(1-F)$. Then $F \in D(G_\gamma)$ if and only if there exists a function $a(t) > 0$ such that*

$$\lim_{t\to+\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \quad \textit{for all } x > 0.$$

Then, a possible choice for the function $a(t)$ is given by

$$a(t) = \begin{cases} \gamma U(t), & \gamma > 0, \\ -\gamma(U(\infty) - U(t)), & \gamma < 0, \\ U(t) - t^{-1}\int_0^t U(s)ds, & \gamma = 0, \end{cases}$$

and a possible choice for the normalization constants in (1) *is*

$$a_m = a(m) \quad \textit{and} \quad b_m = U(m).$$

In the sequel, we will always use the normalization constants (a_m) and (b_m) given in Theorem 1. Note that they are unique up to asymptotic equivalence in the following sense: if (a'_m) and (b'_m) are such that $F^m(a'_m x + b'_m) \to F_\gamma(x)$ for all $x \in \mathbb{R}$, then

$$\lim_{m\to+\infty} \frac{a'_m}{a_m} = 1 \quad \text{and} \lim_{m\to+\infty} \frac{b'_m - b_m}{a_m} = 0. \tag{2}$$

The log-likelihood of the extreme value distribution G_γ is given by

$$\ell_\gamma(x) = -(1 + 1/\gamma)\log(1 + \gamma x) - (1 + \gamma x)^{-1/\gamma},$$

if $1 + \gamma x > 0$ and $-\infty$ otherwise. For $\gamma = 0$, the formula is interpreted as $\ell_0(x) = -x - \exp(-x)$. The three parameter extreme value distribution with shape γ, location μ and scale $\sigma > 0$ has distribution function $x \mapsto F_\gamma(\sigma x + \mu)$. The corresponding log-likelihood is

$$\ell_{(\gamma,\mu,\sigma)}(x) = \ell_\gamma\Big(\frac{x - \mu}{\sigma}\Big) - \log \sigma.$$

The set-up of the block maxima method is the following. We consider independent and identically distributed (i.i.d.) random variables $(X_i)_{i\geq 1}$ with common distribution function $F \in D(G_{\gamma_0})$ and corresponding normalization sequences (a_m) and (b_m) as in Theorem 1. We divide the sequence $(X_i)_{i\geq 1}$ into blocks of length $m \geq 1$ and define the k-th block maximum by

$$M_{k,m} = \max(X_{(k-1)m+1}, \dots, X_{km}), \quad k \geq 1.$$

For fixed $m \geq 1$, the variables $(M_{k,m})_{k\geq 1}$ are i.i.d. with distribution function F^m and

$$\frac{M_{k,m} - b_m}{a_m} \Longrightarrow G_{\gamma_0} \quad \text{as } m \to +\infty. \tag{3}$$

Equation (3) suggests that the distribution of $M_{k,m}$ is approximately a GEV distribution with parameters (γ_0, b_m, a_m) and this is standard to estimate these parameters by the maximum likelihood method. The log-likelihood of the n-sample $(M_{1,m}, \dots, M_{n,m})$ is

$$L_n(\gamma, \sigma, \mu) = \frac{1}{n} \sum_{k=1}^{n} \ell_{(\gamma,\mu,\sigma)}(M_{k,m}).$$

In general, L_n has no global maximum, leading us to the following weak notion: we say that $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE if L_n has a *local* maximum at $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$. Clearly, a MLE solves the likelihood equations

$$\nabla L_n = 0 \quad \text{with} \quad \nabla L_n = \Big(\frac{\partial L_n}{\partial \gamma}, \frac{\partial L_n}{\partial \mu}, \frac{\partial L_n}{\partial \sigma}\Big). \tag{4}$$

Conversely, any solution of the likelihood equations with a definite negative Hessian matrix is a MLE.

For the purpose of asymptotic, we let the length of the blocks $m = m(n)$ depend on the sample size n. Our main result is the following theorem, stating the existence of consistent MLEs.

Theorem 2. *Suppose $F \in D(G_{\gamma_0})$ with $\gamma_0 > -1$ and assume that*

$$\lim_{n\to+\infty} \frac{m(n)}{\log n} = +\infty. \tag{5}$$

Then there exists a sequence of estimators $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ and a random integer $N \geq 1$ such that

$$\mathbb{P}[(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n) \textit{ is a MLE for all } n \geq N] = 1 \tag{6}$$

and

$$\widehat{\gamma}_n \xrightarrow{a.s.} \gamma_0, \quad \frac{\widehat{\mu}_n - b_m}{a_m} \xrightarrow{a.s.} 0 \quad \textit{and} \quad \frac{\widehat{\sigma}_n}{a_m} \xrightarrow{a.s.} 1 \quad \textit{as } n \to +\infty. \tag{7}$$

The condition $\gamma_0 > -1$ is natural and agrees with Smith [17]: it is easy to see that the likelihood equation (4) has no solution with $\gamma \leq 1$ so that no consistent MLE exists when $\gamma_0 < -1$ (see Remark 3 below). Condition (5) states that the block length $m(n)$ grows faster than logarithmically in the sample size n, which is not very restrictive. Let us mention a few further remarks on this condition.

Remark 1. A control of the block size is needed, as the following simple example shows. Consider a distribution $F \in D(G_{\gamma_0})$ with $\gamma_0 > 0$ and such that the left endpoint of F is equal to $-\infty$. Then for each $m \geq 1$, the distribution of the m-th block maxima $M_{k,m}$ has left endpoint equal to $-\infty$ and there exist a sequence $m(n)$ (growing slowly to $+\infty$) such that

$$\lim_{n\to+\infty} \min_{1\leq k\leq n} \frac{M_{k,m} - b_m}{a_m} = -\infty \quad \text{almost surely.} \tag{8}$$

The log-likelihood $L_n(\gamma, \sigma, \mu)$ is finite if and only if

$$\min_{1\leq k\leq n} \Big(1 + \gamma \frac{M_{k,m} - \mu}{\sigma}\Big) > 0,$$

so that any MLE $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ must satisfy

$$\min_{1\leq k\leq n} \Big(1 + \widehat{\gamma}_n \frac{M_{k,m} - \widehat{\mu}_n}{\widehat{\sigma}_n}\Big) > 0.$$

Using this observations, one shows easily that Equation (8) is an obstruction for the consistency (7) of the MLE. Of course, this phenomenon can not happen under condition (5).

Remark 2. It shall be stressed that condition (5) appears only in the proof of Lemma 4 below. One can prove that under stronger assumptions on the distribution $F \in D(G_{\gamma_0})$, condition (5). This is for example the case if F is a Pareto distribution function: one checks easily that the proof of Lemma 4 goes through under the weaker condition $\lim_{n\to+\infty} m(n) = +\infty$. Hence Theorem 2 holds under this weaker condition in the Pareto case. In order to avoid technical conditions that are hard to check in practice when F is unknown, we do not develop this direction any further.

The structure of the paper is the following. We gather in Section 2 some preliminaries on properties of the GEV log-likelihood and of the empirical distribution associated to normalized block maxima. Section 3 is devoted to the proof of Theorem 2, which relies on an adaptation of Wald's method for proving the consistency of M-estimators. Some technical computations (proof of Lemma 4) involving regular variations theory are postponed to an Appendix.

2 Preliminaries

2.1 Properties of the GEV log-likelihood

We gather in the following proposition some basic properties of the GEV log-likelihood. We note x_γ^- and x_γ^+ the left and right end point of the domain ℓ_γ, i.e.

$$(x_\gamma^-, x_\gamma^+) = \{x \in \mathbb{R};\ 1 + \gamma x > 0\}.$$

Clearly, it is equal to $(-\infty, -1/\gamma)$, $\mathbb{R}$ and $(-1/\gamma, +\infty)$ when $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$ respectively.

Proposition 1. *The function ℓ_γ is smooth on its domain.*

1. *If $\gamma \le -1$, ℓ_γ is stricly increasing on its domain and*

$$\lim_{x \to x_\gamma^-} \ell_\gamma(x) = -\infty \qquad \lim_{x \to x_\gamma^+} \ell_\gamma(x) = \begin{cases} +\infty & \text{if } \gamma < -1 \\ 0 & \text{if } \gamma = -1 \end{cases}.$$

2. *If $\gamma > -1$, ℓ_γ is increasing on $(x_\gamma^-, x_\gamma^*]$ and decreasing on $[x_\gamma^*, x_\gamma^+)$, where*

$$x_\gamma^* = \frac{(1+\gamma)^{-\gamma} - 1}{\gamma}.$$

Furthermore

$$\lim_{x \to x_\gamma^-} \ell_\gamma(x) = \lim_{x \to x_\gamma^+} \ell_\gamma(x) = -\infty$$

and ℓ_γ reaches its maximum $\ell_\gamma(x_\gamma^*) = (1+\gamma)(\log(1+\gamma) - 1)$ uniquely.

Remark 3. According to Proposition 1, the log-likelihood ℓ_γ has no local maximum in the case $\gamma \le -1$. This entails that the log-likelihood Equation (4) has no local maximum in $(-\infty, -1] \times \mathbb{R} \times (0, +\infty)$ and that any MLE $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ satisfies $\widehat{\gamma}_n > -1$. Hence, no consistent MLE does exist if $\gamma_0 < -1$. The limit case $\gamma_0 = -1$ is more difficult to analyze and is disregarded in this paper.

2.2 Normalized block maxima

In view of Equation (3), we define the normalized block maxima

$$\widetilde{M}_{k,m} = \frac{M_{k,m} - b_m}{a_m}, \quad k \ge 1,$$

and the corresponding likelihood

$$\widetilde{L}_n(\gamma,\sigma,\mu) = \frac{1}{n}\sum_{k=1}^{n} \ell_{(\gamma,\mu,\sigma)}(\widetilde{M}_{k,m(n)}).$$

It should be stressed that the normalization sequences (a_m) and (b_m) are unknown so that the normalized block maxima $\widetilde{M}_{k,m}$ and the likelihood $\widetilde{L}_n$ cannot be computed from the data only. However, they will be useful in our theoretical analysis since they have good asymptotic properties. The following simple observation will be useful.

Lemma 1. *$(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE if and only if $\widetilde{L}_n$ has a local maximum at $(\widehat{\gamma}_n, (\widehat{\mu}_n - b_m)/a_m, \widehat{\sigma}_n/a_m)$.*

Proof. The GEV likelihood satisfies the scaling property

$$\ell_{\gamma,\mu,\sigma}((x-b)/a) = \ell_{(\gamma,a\mu+b,a\sigma)}(x) + \log a$$

so that

$$L_n(\gamma,\mu,\sigma) = \widetilde{L}_n\Big(\gamma, \frac{\mu-b_m}{a_m}, \frac{\sigma}{a_m}\Big) - \log a_m.$$

Hence the local maximizers of L_n and $\widetilde{L}_n$ are in direct correspondence and the lemma follows. □

2.3 Empirical distributions

The likelihood function $\widetilde{L}_n$ can be seen as a functional of the empirical distribution defined by

$$\mathbb{P}_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{\widetilde{M}_{k,m}},$$

where δ_x denotes the Dirac mass at point $x \in \mathbb{R}$. For any measurable $f : \mathbb{R} \to [-\infty, +\infty)$, we note $\mathbb{P}_n[f]$ the integral with respect to $\mathbb{P}_n$, i.e.

$$\mathbb{P}_n[f] = \frac{1}{n}\sum_{k=1}^{n} f(\widetilde{M}_{k,m}).$$

With these notations, it holds

$$\widetilde{L}_n(\gamma,\mu,\sigma) = \mathbb{P}_n[\ell_{(\gamma,\mu,\sigma)}].$$

The empirical process is defined by

$$\mathbb{F}_n(t) = \mathbb{P}_n((-\infty,t]) = \frac{1}{n}\sum_{k=1}^{n} 1_{\{\widetilde{M}_{k,m}\le t\}}, \quad t \in \mathbb{R}.$$

In the case of an i.i.d. sequence, the Glivenko-Cantelli Theorem states that the empirical process converges almost surely uniformly to the sample distribution function. According to the general theory of empirical processes (see *e.g.* Shorack and Wellner [16] Theorem 1, p106), this result can be extended to triangular arrays of i.i.d. random variables. Equation (3) entails the following result.

Lemma 2. *Suppose* $F \in D(G_{\gamma_0})$ *and* $\lim_{n\to+\infty} m(n) = +\infty$. *Then,*

$$\sup_{t\in\mathbb{R}} |\mathbb{F}_n(t) - F_{\gamma_0}(t)| \xrightarrow{a.s.} 0 \quad \textit{as } n \to +\infty.$$

This entails the almost surely weak convergence $\mathbb{P}_n \Rightarrow G_{\gamma_0}$, whence

$$\mathbb{P}_n[f] \xrightarrow{a.s} G_{\gamma_0}[f] \quad \text{as } n \to +\infty$$

for all bounded and continuous function $f : \mathbb{R} \to \mathbb{R}$. The following lemma dealing with more general functions will be useful.

Lemma 3. *Suppose* $F \in D(G_{\gamma_0})$ *and* $\lim_{n\to+\infty} m(n) = +\infty$. *Then, for all upper semi-continuous function* $f : \mathbb{R} \to [-\infty, +\infty)$ *bounded from above,*

$$\limsup_{n\to+\infty} \mathbb{P}_n[f] \le G_{\gamma_0}[f] \quad \textit{a.s.}.$$

Proof of Lemma 3. Let M be an upper bound for f. The function $\tilde{f} = M - f$ is non-negative and lower semicontinuous. Clearly,

$$\mathbb{P}_n[f] = M - \mathbb{P}_n[\tilde{f}] \quad \text{and} \quad G_{\gamma_0}[f] = M - G_{\gamma_0}[\tilde{f}],$$

whence it is enough to prove that

$$\liminf_{n\to+\infty} \mathbb{P}_n[\tilde{f}] \ge G_{\gamma_0}[\tilde{f}] \quad \text{a.s.}.$$

To see this, we use the relation

$$\mathbb{P}_n[\tilde{f}] = \int_0^1 \tilde{f}(\mathbb{F}_n^{\leftarrow}(u))du.$$

where $\mathbb{F}_n^{\leftarrow}$ is the left-continuous inverse function

$$\mathbb{F}_n^{\leftarrow} = \inf\{x \in \mathbb{R};\ \mathbb{F}_n(x) \ge u\}, \quad u \in (0,1).$$

Lemma 2 together with the continuity of the distribution function F_{γ_0} entail that almost surely, $\mathbb{F}_n^{\leftarrow}(u) \to F_{\gamma_0}^{\leftarrow}(u)$ for all $u \in (0,1)$ as $n \to +\infty$. Using the fact that $\tilde{f}$ is lower semi-continuous, we obtain

$$\liminf_{n\to+\infty} \tilde{f}(\mathbb{F}_n^{\leftarrow}(u)) \ge \tilde{f}(F_{\gamma_0}^{\leftarrow}(u)) \quad u \in (0,1).$$

On the other hand, according to Fatou's lemma,

$$\liminf_{n\to+\infty}\int_0^1 \tilde f(\mathbb{F}_n^{\leftarrow}(u))du \geq \int_0^1 \liminf_{n\to+\infty} \tilde f(\mathbb{F}_n^{\leftarrow}(u))du.$$

Combining the two inequalities, we obtain

$$\liminf_{n\to+\infty} \tilde f(\mathbb{F}_n^{\leftarrow}(u)) \geq \int_0^1 \tilde f(F_{\gamma_0}^{\leftarrow}(u))du,$$

whence

$$\liminf_{n\to+\infty} \mathbb{P}_n[\tilde f] \geq G_{\gamma_0}[\tilde f] \quad \text{a.s..}$$

□

The next lemma plays a crucial role in our proof of Theorem 2. Its proof is quite technical and is postponed to an appendix.

Lemma 4. *Suppose $F \in D(G_{\gamma_0})$ with $\gamma_0 > -1$ and assume condition* (5) *is satisfied. Then,*

$$\lim_{n\to+\infty} \mathbb{P}_n[\ell_{\gamma_0}] = G_{\gamma_0}[\ell_{\gamma_0}] \quad a.s.. \tag{9}$$

It shall be stressed that Lemma 4 is the only part in the proof of Theorem 2 where condition (5) is needed (see Remark 2).

3 Proof of Theorem 2

We introduce the short notation $\Theta = (-1,+\infty)\times\mathbb{R}\times(0,+\infty)$. A generic point of Θ is noted $\theta = (\gamma,\mu,\sigma)$.

The restriction $\widetilde L_n : \Theta \to [-\infty,+\infty)$ is continuous, so that for any compact $K \subset \Theta$, $\widetilde L_n$ is bounded and reaches its maximum on K. We can thus define $\widetilde\theta_n^K = (\widetilde\gamma_n^K, \widetilde\mu_n^K, \widetilde\sigma_n^K)$ such that

$$\widetilde\theta_n^K = \underset{\theta\in K}{\operatorname{argmax}}\ \widetilde L_n(\theta). \tag{10}$$

The following proposition is the key in the proof of Theorem 2.

Proposition 2. *Let $\theta_0 = (\gamma_0,0,1)$ and $K \subset \Theta$ be a compact neighborhood of θ_0. Under the assumptions of Theorem 2,*

$$\lim_{n\to+\infty} \widetilde\theta_n^K = \theta_0 \quad a.s..$$

The proof of Proposition 2 relies on an adaptation of Wald's method for proving the consistency of M-estimators (see Wald [20] or van der Vaart [19] Theorem 5.14). The standard theory of M-estimation is designed for i.i.d. samples, while we have to deal with the triangular array $\{(\widetilde M_{k,m})_{1\leq k\leq n}, n\geq 1\}$. We first state two lemmas.

Lemma 5. *For all $\theta \in \Theta$, $G_{\gamma_0}[\ell_\theta] \leq G_{\gamma_0}[\ell_{\theta_0}]$ and the equality holds if and only if $\theta = \theta_0$.*

Proof of Lemma 5. The quantity $G_{\gamma_0}[\ell_{\theta_0} - \ell_\theta]$ is the Kullback-Leibler divergence of th GEV distributions with parameters θ_0 and θ and is known to be non-negative (see van der Vaart [19] section 5.5). It vanishes if and only if the two distributions agree. This occurs if and only if $\theta = \theta_0$ because the GEV model is identifiable. □

Lemma 6. *For $B \subset \Theta$, define*

$$\ell_B(x) = \sup_{\theta \in B} \ell_\theta(x), \quad x \in \mathbb{R}.$$

Let $\theta \in \Theta$ and $B(\theta, \varepsilon)$ be the open ball in Θ with center θ and radius $\varepsilon > 0$. Then,

$$\lim_{\varepsilon \to 0} G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] = G_{\gamma_0}[\ell_\theta].$$

Proof of Lemma 6. Proposition 1 implies

$$\ell_\theta(x) = \ell_\gamma((x-\mu)/\sigma) - \log \sigma \leq m_\gamma - \log \sigma.$$

One deduce that if B is contained in $(1, \bar{\gamma}] \times [\bar{\sigma}, +\infty) \times \mathbb{R}$ for some $\bar{\gamma} > -1$ and $\bar{\sigma} > 0$, then there exists $M(\bar{\gamma}, \bar{\sigma})$ such that

$$\ell_\theta(x) \leq M(\bar{\gamma}, \bar{\sigma}) \quad \text{for all } \theta \in B, x \in \mathbb{R}.$$

Hence there exists $M > 0$ such that function $M - \ell_{B(\theta,\varepsilon)}$ is non-negative for ε small enough. The continuity of $\theta \mapsto \ell_\theta(x)$ on Θ implies

$$\lim_{\varepsilon \to 0} \ell_{B(\theta,\varepsilon)}(x) = \ell_\theta(x) \quad \text{for all } x \in \mathbb{R}.$$

Then, Fatou's Lemma entails

$$G_{\gamma_0}\big[\liminf_{\varepsilon \to 0}(M - \ell_{B(\theta,\varepsilon)})\big] \leq \liminf_{\varepsilon \to 0} G_{\gamma_0}\big[M - \ell_{B(\theta,\varepsilon)}\big],$$

whence we obtain

$$\limsup_{\varepsilon \to 0} G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] \leq G_{\gamma_0}[\ell_\theta].$$

On the other hand, $\theta \in B(\theta, \varepsilon)$ implies $G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] \geq G_{\gamma_0}[\ell_\theta]$. We deduce

$$\lim_{\varepsilon \to 0} G_{\gamma_0}[\ell_{B(\theta,\varepsilon)}] = G_{\gamma_0}[\ell_\theta].$$

□

Proof of Proposition 2. In view of Lemmas 5 and 6, for each $\theta \in K$ such that $\theta \neq \theta_0$, there exists $\varepsilon_\theta > 0$ such that

$$G_{\gamma_0}[\ell_{B(\theta,\varepsilon_\theta)}] < G_{\gamma_0}[\ell_{\theta_0}].$$

Fix $\delta > 0$. The set $\Delta = \{\theta \in K; \|\theta - \theta_0\| \geq \delta\}$ is compact and is covered by the open balls $\{B(\theta, \varepsilon_\theta), \theta \in \Delta\}$. Let $B_i = B(\theta_i, \varepsilon_{\theta_i})$, $1 \leq i \leq p$, be a finite subcover. Using the relation $\widetilde{L}_n(\theta) = \mathbb{P}_n[\ell_\theta]$, we see that

$$\sup_{\theta\in\Delta} \widetilde{L}_n(\theta) \leq \max_{1\leq i\leq p} \mathbb{P}_n[\ell_{B_i}].$$

The function ℓ_{B_i} is upper semi-continuous and bounded from above, so that Lemma 3 entails

$$\limsup_{n\to+\infty} \mathbb{P}_n[\ell_{B_i}] \leq G_{\gamma_0}[\theta_i] \quad \text{a.s.},$$

whence

$$\limsup_{n\to+\infty} \sup_{\theta\in\Delta} \widetilde{L}_n(\theta) \leq \max_{1\leq i\leq p} G_{\gamma_0}[\theta_i] < G_{\gamma_0}[\theta_0] \quad \text{a.s..} \tag{11}$$

According to Lemma 4, $\mathbb{P}_n[\ell_{\theta_0}] \xrightarrow{a.s.} G_{\gamma_0}[\ell_{\gamma_0}]$, so that

$$\liminf_{n\to+\infty} \sup_{\theta\in K} \widetilde{L}_n(\theta) \geq G_{\gamma_0}[\theta_0] \quad \text{a.s..} \tag{12}$$

Since $\widetilde{\theta}_n^K$ realizes the maximum of $\widetilde{L}_n$ over K, Equations (11) and (12) together entail that $\widetilde{\theta}_n^K \in K\backslash\Delta$ for large n. Equivalently, $\|\widetilde{\theta}_n^K - \theta_0\| < \delta$ for large n. Since δ is arbitrary, this proves the convergence $\widetilde{\theta}_n^K \xrightarrow{a.s.} \theta_0$ as $n \to +\infty$. $\square$

Proof of Theorem 2. Let $K \subset \Theta$ be a compact neighborhood of θ_0 as in Proposition 2 and define $\widetilde{\theta}_n^K$ by Equation (10). We prove that Theorem 2 holds true with the sequence of estimators

$$(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n) = (\widetilde{\gamma}_n^K, a_m\widetilde{\mu}_n^K + b_m, a_m\widetilde{\sigma}_n^K), \quad n \geq 1.$$

According to Lemma 1, $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE if and only if $\widetilde{L}_n$ has a local maximum at $\widetilde{\theta}_n^K = (\widetilde{\gamma}_n^K, \widetilde{\mu}_n^K, \widetilde{\sigma}_n^K)$. Since $\widetilde{\theta}_n^K = \mathrm{argmax}_{\theta\in K} \widetilde{L}_n(\theta)$, this is the case as soon as $\widetilde{L}_n$ lies in the interior set $\mathrm{int}(K)$ of K. Proposition 2 implies the almost surely convergence $\widetilde{\theta}_n^K \xrightarrow{a.s.} \theta_0$ which is equivalent to Equation (7). Furthermore, since $\theta_0 \in \mathrm{int}(K)$, this implies $\widetilde{\theta}_n^K \in \mathrm{int}(K)$ for large n so that $(\widehat{\gamma}_n, \widehat{\mu}_n, \widehat{\sigma}_n)$ is a MLE for large n. This proves Equation (6). $\square$

Acknowledgements

C. Dombry is grateful to Laurens de Haan for suggesting the research problem in the first place and for useful comments that improved greatly an early version of the manuscript.

Appendix: Proof of Lemma 4

We will use the following criterion.

Lemma 7. *Suppose $F \in D(G_{\gamma_0})$ and $\lim_{n\to+\infty} m(n) = +\infty$. We note $Y_m = \ell_{\gamma_0}(a_m^{-1}(M_{1,m} - b_m))$. If there exists a sequence $(\alpha_n)_{n\geq 1}$ and $p > 2$ such that*

$$\sum_{n\geq 1} n\mathbb{P}(|Y_m| > \alpha_n) < +\infty \quad \textit{and} \quad \sup_{n\geq 1} \mathbb{E}[|Y_m|^p 1_{\{|Y_m|\leq \alpha_n\}}] < +\infty,$$

then Equation (9) holds true.

Proof of lemma 7. We note $\mu = G_{\gamma_0}[\ell_{\gamma_0}]$ and we define

$$Y_{k,m} = \ell_{\gamma_0}(a_m^{-1}(M_{k,m} - b_m)) \quad \text{and} \quad S_n = \sum_{k=1}^{n} Y_{k,m}.$$

With these notations, (9) is equivalent to $n^{-1}S_n \overset{a.s.}{\to} \mu$. We introduce the truncated variables

$$\widetilde{Y}_{k,m} = Y_{k,m} 1_{\{|Y_{k,m}|\leq \alpha_n\}} \quad \text{and} \quad \widetilde{S}_n = \sum_{k=1}^{n} \widetilde{Y}_{k,m}.$$

Clearly,

$$\begin{aligned} \mathbb{P}[\widetilde{S}_n \neq S_n] &\leq \mathbb{P}[\widetilde{Y}_{k,m} \neq Y_{k,m} \text{ for some } k \in \{1,\dots,n\}] \\ &\leq n\mathbb{P}[|Y_m| > \alpha_n], \end{aligned}$$

so that $\sum_{n\geq 1} n\mathbb{P}[|Y_m| > \alpha_n] < +\infty$ entails $\sum_{n\geq 1} \mathbb{P}[\widetilde{S}_n \neq S_n] < +\infty$. By the Borel-Cantelli Lemma, this implies that the sequences $(\widetilde{S}_n)_{n\geq 1}$ and $(S_n)_{n\geq 1}$ coincide eventually, whence $n^{-1}S_n \overset{a.s.}{\to} \mu$ if and only if $n^{-1}\widetilde{S}_n \overset{a.s.}{\to} \mu$. We now prove this last convergence.

We first prove that $\mathbb{E}[\widetilde{Y}_{1,m}] \to \mu$. Indeed, by the continuous mapping theorem, the weak convergence (3) implies $Y_{1,m} \Rightarrow \ell_{\gamma_0}(Z)$ with $Z \sim G_{\gamma_0}$. Since $\mathbb{P}[\widetilde{Y}_{1,m} \neq Y_{1,m}]$ converges to 0 as $n \to +\infty$, it also holds $\widetilde{Y}_{1,m} \Rightarrow \ell_{\gamma_0}(Z)$. Together with the condition $\sup_{n\geq 1} \mathbb{E}[|\widetilde{Y}_{1,m}|^p] < \infty$, this entails $\mathbb{E}[\widetilde{Y}_{1,m}] \to \mathbb{E}[\ell_{\gamma_0}(Z)] = \mu$.

Next, Theorem 2.10 in Petrov [12] provides the upper bound

$$\mathbb{E}[|\widetilde{S}_n - \mathbb{E}[\widetilde{S}_n]|^p] \leq C(p)n^{p/2}\mathbb{E}[|\widetilde{Y}_{1,m} - \mathbb{E}[\widetilde{Y}_{1,m}|^p]$$

for some constant $C(p) > 0$ depending only on p. Equivalently,

$$\mathbb{E}[|n^{-1}\widetilde{S}_n - \mu_n|^p] \leq C(p)n^{-p/2}\mathbb{E}[|\widetilde{Y}_{1,m} - \mu_n|^p].$$

with $\mu_n = \mathbb{E}[\widetilde{Y}_{1,m}]$. Furthermore,

$$\mathbb{E}[|\widetilde{Y}_{1,m} - \mu_n|^p] \leq 2^{p-1}(\mathbb{E}[|\widetilde{Y}_{1,m}|^p] + |\mu_n|^p)$$

is uniformly bounded by some constant $C > 0$. By the Markov inequality, for all $\varepsilon > 0$,

$$\mathbb{P}[|n^{-1}\widetilde{S}_n - \mu_n| \geq \varepsilon] \leq \varepsilon^{-p}\mathbb{E}[|n^{-1}\widetilde{S}_n - \mu_n|^p] \leq \varepsilon^{-p}C(p)Cn^{-p/2}.$$

Since $p > 2$, it holds

$$\sum_{n\geq 1}\mathbb{P}[|n^{-1}\widetilde{S}_n - \mu_n| \geq \varepsilon] < +\infty$$

and the Borel-Cantelli Lemma entails $n^{-1}\widetilde{S}_n - \mu_n \overset{a.s.}{\to} 0$. Since $\mu_n \to \mu$, we deduce $n^{-1}\widetilde{S}_n \overset{a.s.}{\to} \mu$ which proves the Lemma. □

Proof of Lemma 4. We prove that there exists a sequence (α_n) and $p > 2$ satisfying

$$\sum_{n\geq 1} n\alpha_n^{-p} < +\infty \tag{13}$$

and

$$\sup_{n\geq 1}\mathbb{E}[(|Y_m| \wedge \alpha_n)^p] < +\infty. \tag{14}$$

The Markov inequality yields

$$\mathbb{P}[|Y_m| \geq \alpha_n] \leq \alpha_n^{-p}\mathbb{E}[(|Y_m| \wedge \alpha_n)^p]$$

so that Equations (13) and (14) together entail

$$\sum_{n\geq 1} n\mathbb{P}(|Y_m| > \alpha_n) < +\infty \quad \text{and} \quad \sup_{n\geq 1}\mathbb{E}[|Y_m|^p 1_{\{|Y_m|\leq\alpha_n\}}] < +\infty,$$

This shows that Equations (13) and (14) together imply the assumptions of Lemma 7 and prove Lemma 4.

We first evaluate the quantity $\mathbb{E}[(|Y_m| \wedge \alpha_n)^p]$ from Equation (14). Recall that $Y_m = \ell_{\gamma_0}((M_{1,m} - b_m)/a_m)$. It is well known that the random variable X_i with distribution function F has the same distribution as the random variable $F^{\leftarrow}(V)$, with V a uniform random variable on

$(0,1)$. We deduce that the random variable $M_{1,m} = \vee_{i=1}^m X_i$ has the same distribution as $F^{\leftarrow}(V_m)$, with V_m a random variable with distribution $mv^{m-1}1_{(0,1)}(v)dv$ (this is the distribution of the maximum of m i.i.d. uniform random variables on $[0,1]$). Hence,

$$\mathbb{E}[(|Y_m| \wedge \alpha_n)^p] = \int_0^1 (|\ell_{\gamma_0}((F^{\leftarrow}(v) - b_m)/a_m)| \wedge \alpha_n)^p m v^{m-1} dv.$$

The relations $U(x) = F^{\leftarrow}(1-1/x)$ and $b_m = U(m)$ together with the change of variable $v = 1 - 1/(mx)$ yield

$$\mathbb{E}[(|Y_m| \wedge \alpha_n)^p] = \int_{1/m}^{\infty} \big(|\ell_{\gamma_0}(\widetilde{U}_m(x))| \wedge \alpha_n\big)^p \big(1 - \frac{1}{mx}\big)^{m-1} x^{-2} dx$$

where

$$\widetilde{U}_m(x) = \frac{U(mx) - U(m)}{a_m}.$$

We now provide an upper bound for the integral and we use the following estimates. There exists a constant $c > 0$ such that

$$|\ell_{\gamma_0}(y)| \leq \begin{cases} c(1+\gamma_0 y)^{-1/\gamma_0}, & y < 0, \\ (1+1/\gamma_0)\log(1+\gamma_0 y) + 1, & y \geq 0 \end{cases}.$$

Note that $\widetilde{U}_m(x)$ is positive for $x > 1$ and negative for $x < 1$. Furthermore, for all $x \geq 1/m$ and $m \geq 2$,

$$\big(1 - \frac{1}{mx}\big)^{m-1} \leq \exp(-(m-1)/(mx)) \leq \exp(-1/(2x)).$$

Using these estimates, we obtain the following upper bound: for $m \geq m_0$ (m_0 to be precised later),

$$\mathbb{E}[(|Y_m| \wedge \alpha_n)^p] \leq I_1 + I_2 + I_3 \tag{15}$$

with

$$\begin{aligned} I_1 &= \int_{1/m}^{m_0/m} \alpha_n^p \exp(-1/(2x)) x^{-2} dx, \\ I_2 &= \int_{m_0/m}^{1} c^p \big(1 + \gamma_0 \widetilde{U}_m(x)\big)^{-p/\gamma_0} \exp(-1/(2x)) x^{-2} dx, \\ I_3 &= \int_1^{\infty} \big((1+1/\gamma_0)\log\big(1+\gamma_0\widetilde{U}_m(x)\big) + 1\big)^p \exp(-1/(2x)) x^{-2} dx. \end{aligned}$$

The integral I_1 can be computed explicitely and

$$I_1 \leq 4\alpha_n^p \exp(-m/(2m_0)). \tag{16}$$

To estimate I_2 and I_3, we need upper and lower bounds for $\widetilde{U}_m(x)$ and we have to distinguish between the three cases $\gamma_0 > 0$, $\gamma_0 \in (-1, 0)$ and $\gamma = 0$.

Case $\gamma_0 > 0$**:** According to Theorem 1, the function U is regularly varying at infinity with index $\gamma_0 > 0$ and

$$1 + \gamma_0 \widetilde{U}_m(x) = 1 + \gamma_0 \frac{U(mx) - U(m)}{a_m} = \frac{U(mx)}{U(m)}.$$

We use then Potter's bound (see e.g. Proposition B.1.9 in [4]): for all $\varepsilon > 0$, there exists $m_0 \geq 1$ such that for $m \geq m_0$ and $mx \geq m_0$

$$(1-\varepsilon)x^{\gamma_0} \min(x^{\varepsilon}, x^{-\varepsilon}) \leq \frac{U(mx)}{U(m)} \leq (1+\varepsilon)x^{\gamma_0} \max(x^{\varepsilon}, x^{-\varepsilon}).$$

We fix $\varepsilon \in (0, \gamma_0)$ and choose m_0 accordingly. Using the lower Potter's bound to estimate I_2 and the upper Potter's bound to estimate I_3, we get

$$\begin{aligned} I_2 &\leq \int_{m_0/m}^{1} c^p \left((1-\varepsilon)x^{\gamma_0+\varepsilon}\right)^{-p/\gamma_0} \exp(-1/(2x))x^{-2}dx \\ &\leq c^p (1-\varepsilon)^{-p/\gamma_0} \int_0^1 x^{-2-p-p\varepsilon/\gamma_0} \exp(-1/(2x))dx, \end{aligned}$$

and

$$I_3 \leq \int_1^{\infty} \left((1+1/\gamma_0)\log\left((1+\varepsilon)x^{\gamma_0+\varepsilon}\right) + 1\right)^p \exp(-1/(2x))x^{-2}dx.$$

These integrals are finite and this implies that I_2 and I_3 are uniformly bounded for $m \geq m_0$. From Equations (15) and (16), we obtain

$$\mathbb{E}[(|Y_m| \wedge \alpha_n)^p] \leq 4\alpha_n^p \exp(-m/(2m_0)) + C,$$

for some constant $C > 0$. Finally, we set $\alpha_n^p \exp(-m/(2m_0)) = 1$, i.e. $\alpha_n = \exp(m/(p2m_0))$. Equation (14) is clearly satisfied and

$$n\alpha_n^{-p} = \exp[\log n - m/(2m_0)] = \exp[-(m/(2m_0 \log n) - 1)\log n].$$

We check easily that the condition $\lim_{n\to+\infty} \frac{m(n)}{\log n} = +\infty$ implies Equation (13).

Case $\gamma_0 < 0$**:** It follows from Theorem 1 that the function $t \mapsto U(\infty) - U(t)$ is regularly varying at infinity with index $\gamma_0 < 0$ and that

$$1 + \gamma_0 \widetilde{U}_m(x) = 1 + \gamma_0 \frac{U(mx) - U(m)}{a_m} = \frac{U(\infty) - U(mx)}{U(\infty) - U(m)}.$$

Then, the Potter's bounds become: for all $\varepsilon > 0$, there exists $m_0 \geq 1$ such that for $m \geq m_0$ and $mx \geq m_0$

$$(1-\varepsilon)x^{\gamma_0}\min(x^{\varepsilon}, x^{-\varepsilon}) \leq \frac{U(\infty)-U(mx)}{U(\infty)-U(m)} \leq (1+\varepsilon)x^{\gamma_0}\max(x^{\varepsilon}, x^{-\varepsilon}).$$

Using this, the proof is completed in the same way as in the case $\gamma_0 > 0$ with straightforward modifications.

Case $\gamma_0 = 0$: In this case, Theorem B.2.18 in [4] implies that for all $\varepsilon > 0$, there exists $m_0 \geq 1$ such that for $m \geq m_0$ and $mx \geq m_0$,

$$\Big|\frac{U(mx)-U(m)}{a_m} - \log x\Big| \leq \varepsilon \max(x^{\varepsilon}, x^{-\varepsilon}).$$

Equivalently, for $m \geq m_0$ and $mx \geq m_0$,

$$\log x - \varepsilon\max(x^{\varepsilon}, x^{-\varepsilon}) \leq \widetilde{U}_m(x) \leq \log x + \varepsilon\max(x^{\varepsilon}, x^{-\varepsilon}).$$

Using the lower bound to estimate I_2 and the upper bound to estimate I_3, we obtain

$$\begin{aligned} I_2 &= \int_{m_0/m}^{1} c^p \exp(-p\widetilde{U}_m(x))\exp(-1/(2x))x^{-2}dx \\ &\leq c^p \int_0^1 \exp(-p\log x + p\varepsilon x^{-\varepsilon} - 1/(2x))x^{-2}dx, \end{aligned}$$

and

$$\begin{aligned} I_3 &= \int_1^{\infty} \big(\widetilde{U}_m(x)+1\big)^p \exp(-1/(2x))x^{-2}dx \\ &\leq \int_1^{\infty} \big(\log x + \varepsilon x^{\varepsilon} + 1\big)^p \exp(-1/(2x))x^{-2}dx. \end{aligned}$$

For $\varepsilon \in (0, 1/p)$, the integrals appearing in the upper bounds are finite and independent of m. This shows that I_2 and I_3 are uniformly bounded for $m \geq m_0$. The proof is then completed as in the case $\gamma_0 > 0$. □

References

[1] A.A. Balkema and L. de Haan. Residual life time at great age. *Ann. Probability*, 2:792–804, 1974.

[2] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of extremes.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.

[3] L. de Haan. Slow variation and characterization of domains of attraction. In *Statistical extremes and applications (Vimeiro, 1983)*, volume 131 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 31–48. Reidel, Dordrecht, 1984.

[4] L. de Haan and A. Ferreira. *Extreme value theory.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.

[5] A.L.M. Dekkers, J.H.J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, 17(4):1833–1855, 1989.

[6] H. Drees, A. Ferreira, and L. de Haan. On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.*, 14(3):1179–1201, 2004.

[7] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, 1997. For insurance and finance.

[8] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5):1163–1174, 1975.

[9] J.R.M. Hosking. Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Applied Statistics*, 34(3):301–310, 1985.

[10] J.R.M. Hosking, J.R. Wallis, and E.F. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261, 1985.

[11] A.J. Macleod. AS R76 - A remark on algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Applied Statistics*, 38(3):198–199, 1989.

[12] V.V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability.* The Clarendon Press Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[13] J. Pickands, III. Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–131, 1975.

[14] P. Prescott and A.T. Walden. Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724, 1980.

[15] P. Prescott and A.T. Walden. Maximum likelihood estimation of the parameters of the three-parameter extreme-value distribution

from censored samples. *Journal of Statistical Computation and Simulation*, 16(3-4):241–250, 1983.

[16] G.R. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[17] R.L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.

[18] R.L. Smith. Estimating tails of probability distributions. *Ann. Statist.*, 15(3):1174–1207, 1987.

[19] A.W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[20] A. Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statistics*, 20:595–601, 1949.

[21] C. Zhou. Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivariate Anal.*, 100(4):794–815, 2009.